

Tilburg University

Phonetic recalibration of speech by text

Keetels, M.N.; Schakel, L.; de Bonte, M.; Vroomen, J.

Published in:
Attention, Perception & Psychophysics

DOI:
[10.3758/s13414-015-1034-y](https://doi.org/10.3758/s13414-015-1034-y)

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Keetels, M. N., Schakel, L., de Bonte, M., & Vroomen, J. (2016). Phonetic recalibration of speech by text. *Attention, Perception & Psychophysics*, 78(3), 938-945. <https://doi.org/10.3758/s13414-015-1034-y>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Phonetic recalibration of speech by text

Mirjam Keetels¹ · Lemmy Schakel¹ · Milene Bonte² · Jean Vroomen¹

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Listeners adjust their phonetic categories to cope with variations in the speech signal (phonetic recalibration). Previous studies have shown that lipread speech (and word knowledge) can adjust the perception of ambiguous speech and can induce phonetic adjustments (Bertelson, Vroomen, & de Gelder in *Psychological Science*, 14(6), 592–597, 2003; Norris, McQueen, & Cutler in *Cognitive Psychology*, 47(2), 204–238, 2003). We examined whether orthographic information (text) also can induce phonetic recalibration. Experiment 1 showed that after exposure to ambiguous speech sounds halfway between /b/ and /d/ that were combined with text (b or d) participants were more likely to categorize auditory-only test sounds in accordance with the exposed letters. Experiment 2 replicated this effect with a very short exposure phase. These results show that listeners adjust their phonetic boundaries in accordance with disambiguating orthographic information and that these adjustments show a rapid build-up.

Keywords Phonetic recalibration · Orthographic information · Rapid recalibration · Letters · Speech perception

Introduction

We are exposed constantly to unclear speech signals due to, for example, differences in dialects, vague pronunciations, bad articulations, or the presence of background noise. To overcome these ambiguities, speech perception needs to be flexible to adjust to these variations. Therefore, perceivers learn that certain speech sounds should be perceived as belonging to a particular speech-sound category. There is ample evidence that lipread speech (and word knowledge) can be used to adjust our perception of ambiguous speech and can induce phonetic adjustments (Bertelson, Vroomen, & de Gelder, 2003; Norris, McQueen, & Cutler, 2003). This phenomenon has now been replicated many times, but to date it has never been examined whether *orthographic* information (i.e., text), can also serve this role. This is of theoretical importance, because text is culturally acquired, whereas lipread speech has strong biological roots with auditory speech perception (Liberman, 1992).

Substantial research has demonstrated that lipreading affects speech perception. One well-studied example is the McGurk effect in which dubbing /gaga/ lip movements on a /baba/ sound typically results in the fused percept of /dada/ (McGurk & MacDonald, 1976). A McGurk aftereffect also has been reported by Bertelson et al. (2003) who showed that repeated exposure to ambiguous speech sounds dubbed onto videos of a face articulating either /aba/ or /ada/ (henceforth: VbA? or VdA? in which Vb = visual /aba/ and A? = auditory ambiguous) induced perceptual adjustments of this ambiguous sound. Thus, after exposure to VbA? an ambiguous sound was perceived as *more* /b/-like than after exposure to VdA?. The underlying notion is that visual lipread input teaches the auditory system how to interpret sounds, a phenomenon called phonetic recalibration. When auditory nonambiguous and congruent exposure stimuli were used (VbAb and VdAd),

✉ Mirjam Keetels
m.n.keetels@uvt.nl

¹ Department of Cognitive Neuroscience and Neuropsychology, Tilburg University, Tilburg, The Netherlands

² Department of Cognitive Neuroscience and Maastricht Brain Imaging Center (M-BIC), Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

no learning effects were observed presumably because there is no intersensory conflict. Rather, aftereffects went in the opposite direction and the ambiguous sound was perceived as *less* /b/-like after exposure to VbAb than VdAd, indicative of “selective speech adaptation” or fatigue (Samuel, 1986; Vroomen, Van Linden, Keetels, De Gelder, & Bertelson, 2004) or a spectral contrast effect (Holt, Lotto, & Kluender, 2000; Holt, Ventura, Rhode, Behesta, & Rinaldo, 2000; Lotto & Kluender, 1998; Lotto, Kluender, & Holt, 1997). Phonetic recalibration by lipread speech has now been replicated many times, also in other laboratories with other tokens and other phonemes. Recent studies show that the phenomenon is phoneme-specific (Reinisch, Wozny, Mitterer, & Holt, 2014) but not speaker-specific (Van der Zande, Jesse, & Cutler, 2014), builds-up fast (Vroomen & Baart, 2009; Vroomen, Van Linden, De Gelder, & Bertelson, 2007), can induce a different interpretation of the same sound simultaneously (Keetels, Pecoraro, & Vroomen, 2015), is comparable for dyslexics (Baart, De Boer-Schellekens, & Vroomen, 2012), and probably involves early cortical auditory networks (Kilian-Hutten, Valente, Vroomen, & Formisano, 2011; Kilian-Hutten, Vroomen, & Formisano, 2011).

Besides lipread-driven phonetic adjustments, *lexical knowledge* can induce phonetic recalibration. Norris et al. (2003) for example demonstrated that exposure to ambiguous sounds embedded in words that normally ended in an /s/ (e.g., *naaldbos*, pine forest) or /f/ (e.g., *witlof*, chicory) resulted in respectively more /s/ or /f/ responses on subsequent ambiguous identification trials (Clarke-Davidson, Luce, & Sawusch, 2008; Eisner & McQueen, 2005; Jesse & McQueen, 2011; Kraljic & Samuel, 2005, 2006, 2007; Myers & Mesite, 2014; Reinisch, Weber, & Mitterer, 2013; Reinisch et al., 2014; Samuel & Kraljic, 2009; Sjerps & McQueen, 2010; Van Linden, Stekelenburg, Tuomainen, & Vroomen, 2007). Van Linden and Vroomen (2007) directly compared lipread and lexically driven recalibration and showed that the effects were comparable in size, build-up, and dissipation rate.

To date, it has never been examined whether these phoneme adjustments also occur when the disambiguating information stems from orthographic information or letters of the alphabet. From an evolutionary point of view, letters are very different from lipread speech or lexical knowledge, because letters are arbitrary cultural artifacts of sound-sight associations that need explicit training during literacy acquisition (Liberman, 1992) whereas for lipread speech there are strong biological constraints between perception and production (Kuhl & Meltzoff, 1982) and the lexicon is acquired in a rather automatic fashion early in life. Therefore, lipread speech and lexical context are both part of the speech signal itself, whereas orthographic information is not, because it only becomes associated with speech during learning to read and occurs

together with speech in specific circumstances, such as reading aloud, subtitles, and psychology experiments.

Despite that letters have different biological roots than lipread speech and the lexicon, there is research that demonstrates that co-occurring letters affect perception of speech sounds. In an early study, Frost, Repp, and Katz (1988) showed that words in noise were identified better when matching text was presented rather than nonmatching text (see also Dijkstra, Schreuder, & Frauenfelder, 1989; Massaro, Cohen, & Thompson, 1988). More recently, it has been reported that acoustically degraded words sound “clearer” if a printed word is seen shortly before the word is heard (Sohoglu, Peelle, Carlyon, & Davis, 2014). The specific brain regions responding to letter-speech congruency also have been studied (Froyen, Van Atteveldt, Bonte, & Blomert, 2008; Sohoglu, Peelle, Carlyon, & Davis, 2012; Van Atteveldt, Formisano, Goebel, & Blomert, 2004). For example, Van Atteveldt et al. (2004) reported that heteromodal regions in the STS as well as early and higher-order auditory cortical regions that are typically involved in speech-sound processing showed letter-speech congruency effects. The authors concluded that the integration of letters and speech sounds relies on a neural mechanism that is similar to the mechanism for integrating lipreading with speech (see also Blau, van Atteveldt, Formisano, Goebel, & Blomert, 2008; Froyen et al., 2008; Van Atteveldt, Formisano, Goebel, & Blomert, 2007). Furthermore, letter-sound congruency effects have been shown to be dependent on reading skills as these effects were less evident in dyslexic readers (Blau et al., 2010; Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009; Blomert, 2011; Froyen, Bonte, Van Atteveldt, & Blomert, 2009; Froyen, Willems, & Blomert, 2011; Kronschnabel, Brem, Maurer, & Brandeis, 2014; Zaric et al., 2014). These studies are indicative of a strong functional coupling between processing of letters and speech sounds.

At present, it is unknown whether letter-speech sound combinations induce aftereffects indicative of phonetic recalibration. To study whether letters do indeed induce phonetic recalibration, we adjusted the original exposure-test paradigm (Bertelson et al., 2003). Participants were exposed to ambiguous (A?) or nonambiguous (Ab or Ad) speech sounds combined with printed text (“aba” or “ada”) and then tested with auditory-only sounds near the phoneme boundary. In Experiment 1, participants were exposed to eight audio-visual exposure stimuli followed by six auditory-only test trials (as in Bertelson et al., 2003). In Experiment 2, the exposure phase was reduced to only one stimulus. If letters acted like lipread speech, we expected that exposure to ambiguous speech sounds combined with disambiguating letters would induce phonetic recalibration.

Experiment 1

Method

Participants Twenty-two students from Tilburg University participated and received course credits for their participation (18 females; 21.7 years average age). Participants reported normal hearing and normal or corrected-to-normal seeing and were fluent Dutch speakers without a diagnosis of dyslexia. They were tested individually and were unaware of the purpose of the experiment. Written, informed consent was obtained from each participant.

Stimuli and materials Participants were seated in front of a 17-inch (600 × 800 pixels) CRT monitor (100-Hz refresh rate) at a distance of approximately 60 cm. The auditory stimuli have been described in detail in Bertelson et al. (2003). In short, we used the audio tracks of a recording of a male Dutch speaker pronouncing the non-words /aba/ and /ada/. The audio was synthesized into a nine-token /aba/-/ada/ continuum by changing the second formant (F2) in eight steps of 39 Mel using the “Praat” speech editor (Boersma & Weenink, 1999). The offset frequency of the first vowel (before the closure) and onset frequency of the second vowel (after the closure) were 1,100 Hz for /aba/ and 1,678 Hz for /ada/ (see Fig. 1 in Vroomen, Van Linden, et al., 2004). The duration of all sound files was 640 ms. From this nine-token continuum, we used the most outer tokens (A1 and A9; henceforth Ab and Ad respectively) and the three middle tokens (A4, A5, and A6; henceforth A?-1, A?, and A?+1, respectively). The audio was delivered binaurally through headphones (Sennheiser HD201) at approximately 66-dB SPL when measured at 5 mm from the earphone.

Visual stimuli consisted of the three letters of the non-words “aba” and “ada” (henceforth Vb and Vd, respectively). The letters were gray (RGB: 128,128,128) presented on a dark background, in the center of the screen (W: 5.5°, H: 2.5°). Visual stimulus duration was 1,200 ms. Letters were presented 450 ms before the sound, because pilot testing showed that this was the most optimal interval to induce perceptual synchrony between the inner speech of the silently read letters (the internal voice that is “heard” while reading) and the externally presented speech sound.

Design and procedure Participants were repeatedly presented with Exposure-Test mini-blocks that each consisted of eight audiovisual exposures followed by six auditory-only test trials. See Fig. 1 for a schematic set-up of the Exposure-Test mini-block design. The exposure stimuli either contained the ambiguous speech sound combined with “aba” or “ada” (VbA? or VdA?) or the nonambiguous speech sound in combination with congruent text (VbAb or VdAd). The interstimulus interval (ISI) between exposure stimuli was 800 ms. The audiovisual exposure phase was followed (after 1,500 ms) by six auditory-only test trials. Test sounds were the most ambiguous token on the continuum (A?), its more “aba-like” neighbor (A?-1), and the more “ada-like” neighbor on the continuum (A?+1). The three test sounds (A?-1; A?; A?+1) were presented twice in random order. The participant’s task was to indicate whether the test sound was /aba/ or /ada/ by pressing a corresponding key on a response box. The intertrial interval (ITI) was 1,250 ms.

Each participant completed 36 Exposure-Test miniblocks where each of the 4 exposure stimuli was presented 9 times (to collect 18 repetitions of each test sound per exposure condition). There was a short pause after 12 mini-blocks. The

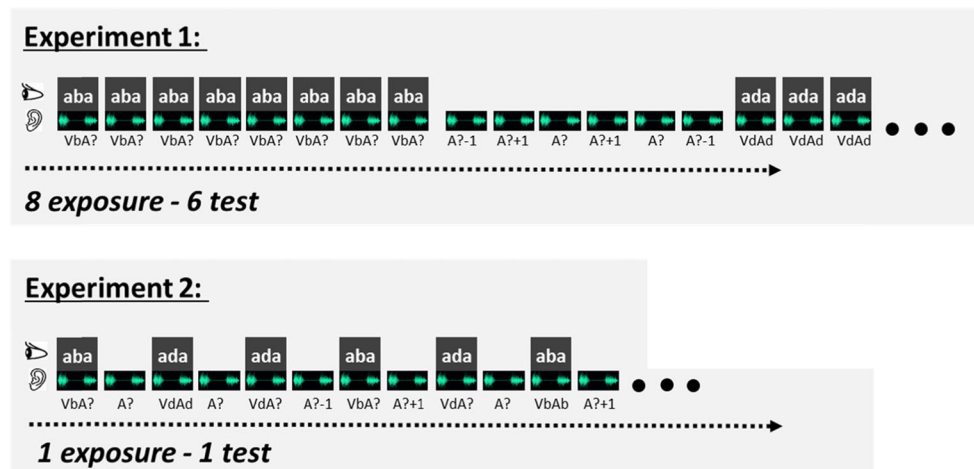


Fig. 1 Schematic overview of the Exposure-Test paradigm. In Experiment 1, participants were exposed to 8 auditory-visual exposure stimuli followed by 6 auditory-only test trials (i.e., 8 exposure - 6 test). In Experiment 2,

participants were exposed to 1 auditory-visual exposure stimulus followed by 1 auditory-only test trial (i.e., 1 exposure - 1 test).

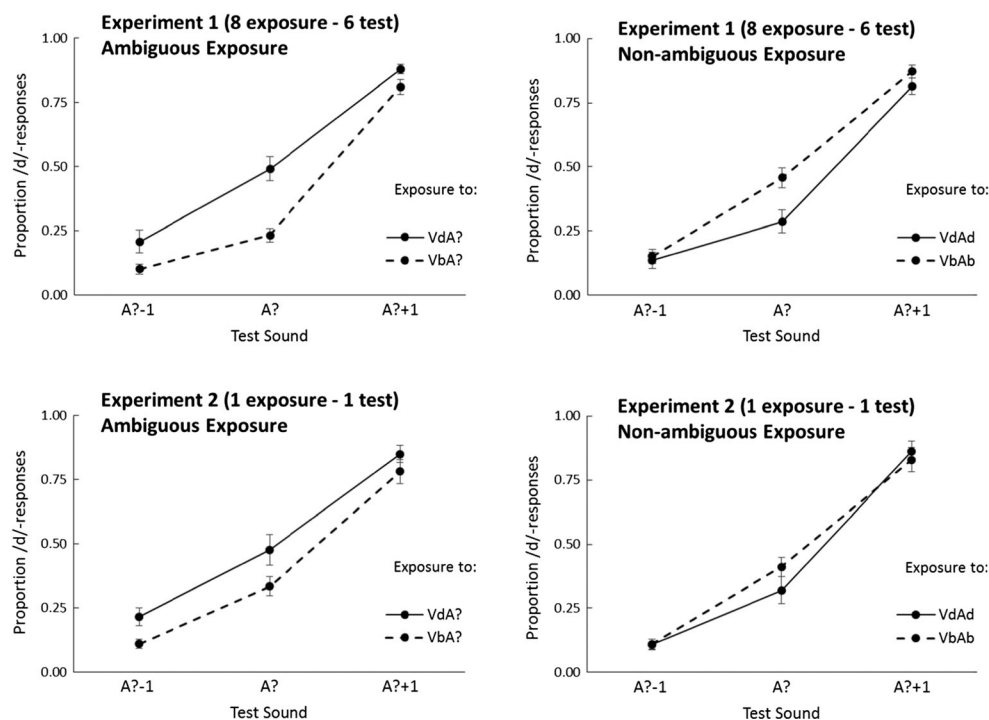


Fig. 2 Proportion /d/ responses for the three different test-sounds (A?-1, A?, and A?+1) after ambiguous exposure to VbA? or VdA? (left panel) and nonambiguous exposure to VbAb and VdAd (right panel) for

Experiment 1 (upper panels) and Experiment 2 (lower panels). Error bars represent the standard errors of the mean.

audiovisual exposure stimuli varied randomly between mini-blocks.

Results

The individual proportion of /d/ responses on the auditory-only test trials was calculated for each combination of exposure-sound (ambiguous or nonambiguous), exposure text (Vb or Vd), and test sound (A?-1; A?; A?+1). Figure 2 displays the average proportions of /d/ responses. Most importantly, for ambiguous exposure sounds, there were *more* /d/ responses after exposure to VdA? than after VbA? (Vd-Vb difference = 0.15; indicative of phonetic recalibration). For nonambiguous exposure, there were *fewer* /d/ responses after exposure to VdAd than after VbAb (Vd-Vb difference = -0.08, indicative of selective speech adaptation or spectral contrast effect).

This was confirmed in a generalized linear mixed-effects model with a logistic linking function to account for the dichotomous dependent variable (lme4 package in R version 3.2.2). The dependent variable *Response* was recoded in such a way that a /b/-response was coded as “0” and a /d/-response was coded as “1” (therefore a positive-fitted coefficient reflects more /d/ responses). The factor *Exposure-sound* was recoded into +0.5 for ambiguous, and -0.5 for nonambiguous (therefore the fitted coefficient will correspond to the difference between nonambiguous and ambiguous conditions). Similarly, the factor *Exposure-text* was recoded into +0.5 for

Vb, and -0.5 for Vd. The factor *Test-sound* was entered as a numeric factor centered on zero (-1 for A?-1; 0 for A?; and +1 for A?+1) such that the fitted coefficient will correspond to the slope of the /b/-d/ classification boundary (in units of change in log-odds of /d/-response per one continuum step).

The fitted model included *Response* (/b/ or /d/-response) as the dependent variable, and *Exposure-sound* (ambiguous or nonambiguous), *Exposure-text* (Vb or Vd), *Test-sound* (A?-1; A?; A?+1), and their interactions as fixed factors (see Table 1). The fitted model was: $\text{Response} \sim 1 + \text{Exposure-sound} * \text{Exposure-text} * \text{Test-sound} + (1 + \text{Exposure-text} + \text{Exposure-sound} : \text{Exposure-text} + \text{Test-sound} \parallel \text{Subject})^1$. All the fixed effects correlations were less than 0.2.

The analysis revealed a significant negative effect for the intercept ($b = -0.42$, $SE = 0.14$, $p < 0.01$), which indicates a slight /b/-bias overall. There was no main effect of *Exposure-text* ($b = -0.17$, $SE = 0.15$, $p = 0.29$) but a significant main effect of *Test-sound* ($b = 2.28$, $SE = 0.21$, $p < 0.001$) indicative of more /d/ responses for the more /d/-like test sounds. Importantly, a significant interaction between *Exposure-sound* and

¹ We used the maximal random effect structure supported by the data (uncorrelated random effects by subjects for intercepts and slopes for *Exposure-text*, *Test-sound*, and the *Exposure-sound*-by-*Exposure-text* interaction). We determined this random effects structure by starting with a maximal model and removing slopes for fixed effects we are not directly interested in (i.e., random effect correlations, the effect of *Exposure-sound* and interactions between *Test-sound* and *Exposure-sound*/*Exposure-text*) until the model converged (Bates, Kliegl, Vasishth, *in press*).

Table 1 Fitted model: Response $\sim 1 + \text{Exposure-sound} * \text{Exposure-text} * \text{Test-sound} + (1 + \text{Exposure-text} + \text{Exposure-sound}:\text{Exposure-text} + \text{Test-sound} \parallel \text{Subject})$

| Experiment | Fixed factor | Estimate | Standard error | z-value | p |
|------------|---|----------|----------------|---------|-----------|
| 1 | (Intercept) | -0.42 | 0.14 | -2.92 | <0.01** |
| | Exposure-sound | -0.01 | 0.08 | -0.10 | 0.92 |
| | Exposure-text | -0.17 | 0.15 | -1.07 | 0.29 |
| | Test-sound | 2.28 | 0.21 | 11.09 | <0.001*** |
| | Exposure-sound*Exposure-text | -1.73 | 0.23 | -7.50 | <0.001*** |
| | Exposure-sound*Test-sound | -0.01 | 0.11 | -0.08 | 0.93 |
| | Exposure-text*Test-sound | 0.17 | 0.11 | 1.46 | 0.14 |
| | Exposure-sound*Exposure-text*Test-sound | 0.36 | 0.23 | 1.58 | 0.11 |
| 2 | (Intercept) | -0.45 | 0.21 | -2.12 | 0.03* |
| | Exposure-sound | 0.21 | 0.10 | 2.08 | 0.04* |
| | Exposure-text | -0.22 | 0.25 | -0.88 | 0.38 |
| | Test-Sound | 2.57 | 0.23 | 11.31 | <0.001*** |
| | Exposure-sound*Exposure-text | -1.04 | 0.26 | -4.04 | <0.001*** |
| | Exposure-sound*Test-sound | -0.50 | 0.15 | -3.41 | <0.001*** |
| | Exposure-text*Test-sound | -0.16 | 0.16 | -1.01 | 0.31 |
| | Exposure-sound*Exposure-text*Test-sound | 0.56 | 0.30 | 1.87 | 0.06 |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Exposure-text was found ($b = -1.73$, $SE = 0.23$, $p < 0.001$), indicating that aftereffects (i.e., Vd-Vb difference) were different for ambiguous and nonambiguous exposure sounds. Significant or nonsignificant effects of any within-subject variables that did not have random slopes are not interpreted.

To further explore the interaction effect between Exposure-sound and Exposure-text, Bonferroni corrected pairwise contrasts were performed and showed a higher proportion of /d/ responses after exposure to VdA? compared with VbA? (data collapsed for the three Test-sounds, $b = 1.03$, $SE = 0.19$, $p < 0.001$). For the nonambiguous Exposure sound, higher proportions of /d/ responses after exposure to VbAb compared with VdAd ($b = -0.70$, $SE = 0.19$, $p < 0.001$). The results of Experiment 1 demonstrate that phonetic recalibration can be induced by orthographic information.

Experiment 2

In Experiment 2, we further examined whether this effect also can be induced by a very short exposure phase. Evidence for rapid recalibration has been reported before for phonetic recalibration by lipread speech (Vroomen et al., 2007) and for temporal and spatial recalibration (Harvey, Van der Burg, & Alais, 2014; Van der Burg, Alais, & Cass, 2013; Van der Burg & Goodbourn, 2015). Temporal recalibration refers to the phenomenon in which exposure to a certain temporal relation between a sound and a light (i.e., sound-before-light or light-before-sound) results in adjustments of perceived intersensory

timing (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, De Gelder, & Bertelson, 2004). Van der Burg et al. (2013) showed that exposure to only a single asynchronous auditory-visual exposure stimulus is sufficient to induce strong temporal recalibration effects afterwards. This led the authors to conclude that temporal recalibration is a fast-acting process that serves to rapidly realign sensory signals (see also Harvey et al., 2014). Similar effects have been reported in the spatial domain by Wozny and Shams (2011), showing that recalibration of perceived auditory space by vision can occur after a single exposure to discrepant auditory-visual stimuli lasting only a few milliseconds. Based on these findings, we hypothesized that phonetic recalibration by orthographic information also might show a rapid build-up. To examine this, participants were exposed to a single audiovisual exposure stimulus followed by an auditory-only test-trial.

Methods

Twenty-two participants were tested (18 females; 20.1 years average age). Stimuli, procedure, and design were as in Experiment 1, except that an Exposure-Test mini-block consisted of a single exposure-trial (VbA?, VdA?, VbAb, or VdAd) followed by one test-trial (A?-1, A?, or A?+1, randomly varied between mini-blocks). Figure 1 (lower panel) shows the schematic set-up of an Exposure-Test mini-block. Each participant completed 144 Exposure-Test mini-blocks with a short pause after 72 mini-blocks (to collect 12 repetitions per condition).

Results

After exposure to the ambiguous sounds, there were *more* /d/ responses after exposure to VdA? than after VbA? ($Vd-Vb = 0.10$, indicative of fast phonetic recalibration), whereas for nonambiguous exposure stimuli there was no difference ($Vd-Vb = -0.02$).

Data were analyzed as in Experiment 1. The generalized linear mixed-effects model ($\text{Response} \sim 1 + \text{Exposure-sound} * \text{Exposure-text} * \text{Test-sound} + (1 + \text{Exposure-text} + \text{Exposure-sound} : \text{Exposure-text} + \text{Test-sound} \parallel \text{Subject})$) revealed a significant negative effect for the intercept ($b = -0.45$, $SE = 0.21$, $p < 0.05$) which indicates a slight /b/-bias overall. There was no main effect of Exposure-text ($b = -0.22$, $SE = 0.25$, $p = 0.38$) but a significant main effect of Test-sound ($b = 2.57$, $SE = 0.23$, $p < 0.001$) indicative of more /d/ responses for the more d-like test sounds. Importantly, a significant interaction between Exposure-sound and Exposure-text was found ($b = -1.04$, $SE = 0.26$, $p < 0.001$), indicating that aftereffects (i.e., Vd-Vb difference) were different for ambiguous and nonambiguous exposure sounds. Significant or nonsignificant effects of any within-subject variables that did not have random slopes are not interpreted.

To further explore the interaction effect between Exposure-sound and Exposure-text, Bonferroni corrected pairwise contrasts were performed and showed a higher proportion of /d/ responses after exposure to VdA? compared with VbA? (data collapsed for the three Test-sounds ($b = 0.74$, $SE = 0.28$, $p < 0.01$)). For the nonambiguous Exposure sound, no difference in proportions of /d/ responses were found between exposure to VbAb and VdAd ($b = -0.30$, $SE = 0.28$, $p = 0.29$).

General discussion

The phoneme boundary between two speech categories is flexible and earlier studies have shown that this boundary can be readjusted by lipread speech or lexical word knowledge that tells what the sound should be (Bertelson et al., 2003; Norris et al., 2003). We report, for the first time, that *orthographic information* (text) also can serve this role. Phonetic recalibration can be induced by letter-speech sound combinations. These adjustments show a fast build-up, even after a single letter-sound exposure. These sound-sight associations that are culturally defined and acquired by extensive reading training can thus adjust auditory perception at the phoneme level.

How does this finding of orthographically induced phonetic recalibration relate to lipread and lexically induced phonetic recalibration? Although the three types of inducer stimuli seem rather different in nature and magnitudes of lipread-induced phonetic recalibration effects seem somewhat bigger (~20–40 % in Bertelson et al., 2003 and Van Linden and Vroomen, 2007) compared with the effects reported in the

present study (15 % in Experiment 1 and 10 % in Experiment 2), they appear to rely on a common underlying factor. Van Linden and Vroomen (2007) showed similar characteristics for lipread and lexical-induced recalibration and concluded that the lipread and lexical information serve the same role in phonetic adjustments. The findings of the present study demonstrate that orthographical information also can serve this disambiguating role to induce phonetic adjustments.

We also demonstrated that orthographic phonetic recalibration builds-up very quickly, because it is already stable after a single-exposure stimulus. This finding is in line with previous reports on rapid build-up of phonetic recalibration by lipread stimuli (Vroomen et al., 2007). For temporal and spatial recalibration, a similar fast build-up has been found (Harvey et al., 2014; Van der Burg et al., 2013; Wozny & Shams, 2011). It seems safe to conclude that recalibration behaves similarly such that discrepancies between the senses—either in space, time, or phonetic identity—are rapidly minimized by adjustments of the unreliable source.

Would orthographically induced recalibration be affected when letter-speech sound binding is impaired? Recent studies have reported that impairments in the automatic integration of letters and speech sounds is associated with dyslexia (Blau et al., 2010; Blau et al., 2009; Froyen et al., 2011; Kronschnabel et al., 2014; Zarić et al., 2014; Žarić et al., 2015). Froyen et al. (2011) for example showed in a mismatch-negativity (MMN) paradigm that dyslexics do not exhibit the typical early influences of letters on speech sounds, despite several years of reading instruction. Given that dyslexia has been linked to impairments in grapho-phonological conversions, it might be the case that dyslexic readers do not show phonetic recalibration as normal readers do. This may be an interesting contrast with lipread-induced recalibration, which appears to be in the normal range in people with dyslexia (Baart et al., 2012).

The finding that exposure to nonambiguous speech sounds did not evoke contrast effects in Experiment 2 may appear in conflict with previous findings on selective speech adaptation (Eimas & Corbit, 1973; Samuel, 1986; Vroomen, Keetels, et al., 2004). This null-effect can be explained by the length of the exposure phase, given that selective speech adaptation requires more extensive amounts of exposure.

To summarize, the present study demonstrates that phonetic recalibration can be induced by text very rapidly. Together with previous findings, this is evidence that different information sources (lipread speech, lexical information, text), whether biologically rooted in speech or culturally acquired, can all gain access to the phonetic system.

Acknowledgments The authors thank Dave Kleinschmidt for assistance with adopting the generalized linear mixed-effects model analysis.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Baart, M., De Boer-Schellekens, L., & Vroomen, J. (2012). Lipread-induced phonetic recalibration in dyslexia. *Acta Psychologica*, 140(1), 91–95.
- Bates D., Kliegl R., Vasishth S., & Baayen, H. (in press). Parsimonious mixed models. *Journal of Memory and Language*.
- Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, 14(6), 592–597. doi:10.1046/j.0956-7976.2003.psci.1470.x
- Blau, V., Reithler, J., van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., & Blomert, L. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: A functional magnetic resonance imaging study of dyslexic children. *Brain*, 133, 868–879. doi:10.1093/brain/awp308
- Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology*, 19(6), 503–508. doi:10.1016/j.cub.2009.01.065
- Blau, V., van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2008). Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *European Journal of Neuroscience*, 28(3), 500–509. doi:10.1111/j.1460-9568.2008.06350.x
- Blomert, L. (2011). The neural signature of orthographic–phonological binding in successful and failing reading development. *NeuroImage*, 57(3), 695–703.
- Boersma, P., & Weenink, D. (1999). PRAAT: Doing phonetics by computer. *Software*.
- Clarke-Davidson, C., Luce, P., & Sawusch, J. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, 70(4), 604–618. doi:10.3758/pp.70.4.604
- Dijkstra, T., Schreuder, R., & Frauenfelder, U. H. (1989). Grapheme context effects on phonemic processing. *Language and Speech*, 32(2), 89–108.
- Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4(1), 99–109.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67(2), 224–238. doi:10.3758/bf03206487
- Frost, R., Repp, B. H., & Katz, L. (1988). Can speech-perception be influenced by simultaneous presentation of print. *Journal of Memory and Language*, 27(6), 741–755. doi:10.1016/0749-596x(88)90018-6
- Froyen, D., Bonte, M., Van Atteveldt, N., & Blomert, L. (2009). The long road to automation: Neurocognitive development of letter-speech sound processing. *Journal of Cognitive Neuroscience*, 21(3), 567–580. doi:10.1162/jocn.2009.21061
- Froyen, D., Van Atteveldt, N., Bonte, M., & Blomert, L. (2008). Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neuroscience Letters*, 430(1), 23–28. doi:10.1016/j.neulet.2007.10.014
- Froyen, D., Willems, G., & Blomert, L. (2011). Evidence for a specific cross-modal association deficit in dyslexia: An electrophysiological study of letter-speech sound processing. *Developmental Science*, 14(4), 635–648.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, 7(7), 773–778. doi:10.1038/nn1268
- Harvey, C., Van der Burg, E., & Alais, D. (2014). Rapid temporal recalibration occurs crossmodally without stimulus specificity but is absent unimodally. *Brain Research*, 1585, 120–130.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000a). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, 108(2), 710–722. doi:10.1121/1.429604
- Holt, L. L., Ventura, V., Rhode, W. R., Behesta, S., & Rinaldo, A. (2000b). Context-dependent neural coding in the chinchilla cochlear nucleus. *Journal of the Acoustical Society of America*, 108, 2641.
- Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, 18(5), 943–950.
- Keetels, M., Pecoraro, M., & Vroomen, J. (2015). Recalibration of auditory phonemes by lipread speech is ear-specific. *Cognition*, 141, 121–126.
- Kilian-Hutten, N., Valente, G., Vroomen, J., & Formisano, E. (2011a). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *Journal of Neuroscience*, 31(5), 1715–1720. doi:10.1523/jneurosci.4572-10.2011
- Kilian-Hutten, N., Vroomen, J., & Formisano, E. (2011b). Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. *NeuroImage*, 57(4), 1601–1607. doi:10.1016/j.neuroimage.2011.05.043
- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2), 141–178.
- Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, 13(2), 262–268.
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15.
- Kronshabel, J., Brem, S., Maurer, U., & Brandeis, D. (2014). The level of audiovisual print–speech integration deficits in dyslexia. *Neuropsychologia*, 62, 245–261.
- Kuhl, P., & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138–1141. doi:10.1126/science.7146899
- Liberman, A. M. (1992). The relation of speech to reading and writing. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology and meaning* (pp. 167–178). Amsterdam, The Netherlands: Elsevier Science Publishers BV.
- Lotto, A. J., & Kluender, K. R. (1998). General auditory processes may account for the effect of preceding liquid on perception of place of articulation. *Perception & Psychophysics*, 60, 602–619.
- Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, 102(2), 1134–1140. doi:10.1121/1.419865
- Massaro, D. W., Cohen, M. M., & Thompson, L. A. (1988). Visible language in speech-perception, lipreading and reading. *Visible Language*, 22(1), 9–31.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748. doi:10.1038/264746a0
- Myers, E., & Mesite, L. (2014). Neural systems underlying perceptual adjustment to non-standard speech tokens. *Journal of Memory and Language*, 76, 80–93. doi:10.1016/j.jml.2014.06.007
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238. doi:10.1016/s0010-0285(03)00006-9
- Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 75–86. doi:10.1037/a0027979

- Reinisch, E., Wozny, D., Mitterer, H., & Holt, L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, 45, 91–105. doi:10.1016/j.wocn.2014.04.002
- Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology*, 18(4), 452–499.
- Samuel, A., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71(6), 1207–1218. doi:10.3758/app.71.6.1207
- Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 195–211. doi:10.1037/a0016803
- Sohoglu, E., Peelle, J., Carlyon, R., & Davis, M. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, 32(25), 8443–8453. doi:10.1523/Jneurosci.5069-11.2012
- Sohoglu, E., Peelle, J., Carlyon, R., & Davis, M. (2014). Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 186–199. doi:10.1037/a0033206
- Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43(2), 271–282.
- Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2007). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *NeuroImage*, 36(4), 1345–1360. doi:10.1016/j.neuroimage.2007.03.065
- Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *The Journal of Neuroscience*, 33(37), 14633–14637.
- Van der Burg, E., & Goodbourn, P. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proceedings of the Royal Society B: Biological Sciences*, 282(1804), 8. doi:10.1098/rspb.2014.3083
- Van der Zande, P., Jesse, A., & Cutler, A. (2014). Cross-speaker generalisation in two phoneme-level perceptual adaptation processes. *Journal of Phonetics*, 43, 38–46.
- Van Linden, S., Stekelenburg, J. J., Tuomainen, J., & Vroomen, J. (2007). Lexical effects on auditory speech perception: An electrophysiological study. *Neuroscience Letters*, 420(1), 49–52.
- Van Linden, S., & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1483.
- Vroomen, J., & Baart, M. (2009). Recalibration of phonetic categories by lipread speech: Measuring aftereffects after a 24-hour delay. *Language and Speech*, 52, 341–350. doi:10.1177/0023830909103178
- Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004a). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, 22(1), 32–35. doi:10.1016/j.cogbrainres.2004.07.003
- Vroomen, J., Van Linden, S., Keetels, M., De Gelder, B., & Bertelson, P. (2004b). Selective adaptation and recalibration of auditory speech by lipread information: Dissipation. *Speech Communication*, 44(1), 55–61.
- Vroomen, J., Van Linden, S., De Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia*, 45(3), 572–577. doi:10.1016/j.neuropsychologia.2006.01.031
- Wozny, D., & Shams, L. (2011). Recalibration of auditory space following milliseconds of cross-modal discrepancy. *Journal of Neuroscience*, 31(12), 4607–4612. doi:10.1523/jneurosci.6079-10.2011
- Zaric, G., Gonzalez, G., Tijms, J., van der Molen, M., Blomert, L., & Bonte, M. (2014). Reduced neural integration of letters and speech sounds in dyslexic children scales with individual differences in reading fluency. *Plos One*, 9(10). doi:10.1371/journal.pone.0110337
- Žaric, G., González, G., Tijms, J., van der Molen, M., Blomert, L., & Bonte, M. (2015). Crossmodal deficit in dyslexic children: practice affects the neural timing of letter-speech sound integration. *Frontiers in Human Neuroscience*, 9. doi:10.3389/fnhum.2015.00369